Test-cost-sensitive attribute reduction of data with normal distribution measurement errors

Hong Zhao, Fan Min*, William Zhu

Lab of Granular Computing, Zhangzhou Normal University, Zhangzhou 363000, China

Abstract

The measurement error with normal distribution is universal in applications. Generally, smaller measurement error requires better instrument and higher test cost. In decision making based on attribute values of objects, we shall select an attribute subset with appropriate measurement error to minimize the total test cost. Recently, error-range-based covering rough set with uniform distribution error was proposed to investigate this issue. However, the measurement errors satisfy normal distribution instead of uniform distribution which is rather simple for most applications. In this paper, we introduce normal distribution measurement errors to covering-based rough set model, and deal with test-cost-sensitive attribute reduction problem in this new model. The major contributions of this paper are four-fold. First, we build a new data model based on normal distribution measurement errors. With the new data model, the error range is an ellipse in a two-dimension space. Second, the covering-based rough set with normal distribution measurement errors is constructed through the "3-sigma" rule. Third, the test-cost-sensitive attribute reduction problem is redefined on this covering-based rough set. Fourth, a heuristic algorithm is proposed to deal with this problem. The algorithm is tested on ten UCI (University of California - Irvine) datasets. The experimental results show that the algorithm is more effective and efficient than the existing one. This study is a step toward realistic applications of cost-sensitive learning.

Keywords: Normal distribution, measurement errors, test costs, covering-based rough set.

1. Introduction

The measurement error is the difference between a measurement value and its true value. It can come from the measuring instrument, from the item being measured, from the environment, from the operator, and from other sources

*Corresponding author. Tel.: +86 133 7690 8359

Email addresses: hongzhao2012@163.com (Hong Zhao), minfanphd@163.com (Fan Min), williamfengzhu@gmail.com (William Zhu)

[1]. As a plausible distribution for measurement errors, the normal distribution was put forward by Gauss in 1809. In fact, normal distribution is found to be applicable over almost the whole of science and engineering measurement. In data mining applications, the data model based on measurement errors is an important form of uncertain data (see, e.g., [2, 3, 4]).

Test costs refer to time, money, or other resources spent in obtaining data items related to some object [5, 6, 7, 8, 9, 10]. There are a number of measurement methods with different test costs to obtain a data item. Generally, higher test cost is required to obtain data with smaller measurement error. In data mining applications, we shall select an attribute subset with appropriate measurement error to minimize the total test cost, and at the same time preserve necessary information of the original decision system.

An attribute reduct is a subset of attributes that are jointly sufficient and individually necessary for preserving a particular property of the given information table [11]. It is a key problem of rough set theory and has attracted much attention in recent years (see, e.g., [12, 13, 14, 15, 16]). As a generalization of attribute reduction, test-cost-sensitive attribute reduction [9] focuses on selecting a set of tests to satisfy a minimal test cost criterion.

Recently, error-range-based covering rough set [4] was introduced to address error ranges. This theory is based on both covering-based rough set [17, 18, 19, 20, 21, 22, 23] and neighborhood rough set [24, 25, 26, 27, 28]. Moreover, in the new theory, the test-cost-sensitive attribute reduction problem deals with numeric data instead of nominal ones. Therefore the problem is more challenging than the one defined in [9]. However, error-range-based covering rough set considers only uniform distribution errors, which are rather unrealistic.

In this paper, we introduce normal distribution to build a new model of covering-based rough set to address measurement errors (NDME) according to the "3-sigma" rule. The major contributions of this paper are four-fold. First, we introduce normal distribution to build a new data model based on measurement errors. With the new data model, the error range is an ellipse in a two-dimension space. The error range is computed according to the values of attributes instead of the fixed error range for different datasets. Second, we build the computational model, namely the covering-based rough set with normal distribution measurement errors. Third, the minimal test cost attribute reduction problem is redefined in the new model. Fourth, we propose a heuristic algorithm to address the reduction problem. Specifically, a δ-weighted heuristic reduction algorithm is designed, where attribute significance is adjusted by δ-weighted test cost.

Ten open datasets from the UCI library are employed to study the performance and effectiveness of our algorithm. We adopt three metrics to evaluate the performance of the reduction algorithms from a statistical viewpoint. Experiments undertaken with open source software Coser [29] validate the performance of this algorithm. Experimental results show that our algorithm can generate a minimal test cost reduct in most cases. At the same time, the proposed algorithm can achieve better performance and efficiency than the existing one [4].

The rest of the paper is organized as follows: Section 2 presents the data models with measurement errors and test costs, respectively. Section 3 describes the computational model, namely covering-based rough set model with normal distribution measurement errors. The minimal test cost reduct problem under the new model is also defined in this section. Next, Section 4 presents a δ-weighted heuristic reduction algorithm and a competition approach. Experiment results and comparison with the existing work are discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. Data models

This section presents data models. First, we propose a decision system with normal distribution measurement errors, which is also called NEDS for brevity. Then, we introduce test costs to NEDS, and define test-cost-sensitive decision systems with NDME.

2.1. Normal distribution measurement errors

Normal distribution is symmetrical with a single central peak at the mean of the data [30]. It is described by the probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{1}$$

where parameters μ is the mean and σ^2 is the variance.

The cumulative distribution function $F(x)$ describes probability of a random variable falling in the interval $(-\infty, x]$.

$$F(x) = \int_{-\infty}^{x} f(x)dx, \tag{2}$$

where $x \in \mathbb{R}$.

For a random variable X,

$$Pr(X \leq x) = F(x). \tag{3}$$

The standard normal distribution appears with $\mu = 0$ and $\sigma^2 = 1$. The equation becomes

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}. \tag{4}$$

As shown in Figure 1, if your data obey a normal distribution, over 99% of your subjects will fall within three standard deviations of the mean. We use the following example to explain the relationship between standard deviation and confidence interval.

Example 1. *Let standard deviation be 0.01, the mean be 0, then we know that about 99% of the measurement errors from -0.03 to +0.03.*

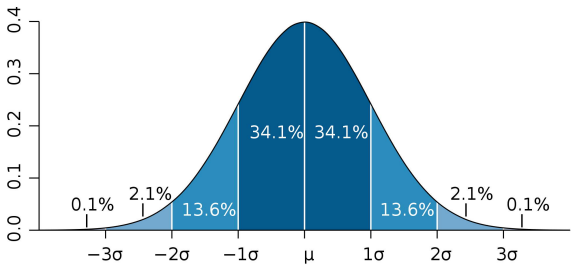

Figure 1: Confidence interval.

2.2. Decision systems with normal distribution measurement errors

We introduce normal distribution measurement errors into our model to expand the application scope. For a normal distribution, nearly all values lie within 3 standard deviations of the mean, that is "3-sigma" rule [30].

Definition 2. A *decision system with normal distribution measurement errors* (NEDS) S is the 6-tuple:

$$\begin{aligned} S &= (U, C, D, V = \{V_a | a \in C \cup D\}, \\ &\quad I = \{I_a | a \in C \cup D\}, n), \end{aligned} \tag{5}$$

where U is the nonempty set called a universe, C and D are the nonempty sets of variables called as conditional attributes and decision attributes, respectively. V_a is the set of values for each $a \in C \cup D$, and $I_a : U \to V_a$ is an information function for each $a \in C \cup D$. We often denote $\{V_a | a \in C \cup D\}$ and $\{I_a | a \in C \cup D\}$ by V and I, respectively. $n : C \to \mathbb{R}^+ \cup \{0\}$ is the maximum value of measurement error. $+n(a)$ and $-n(a)$ are the upper confidence limit(UCL) and the lower confidence limit(LCL) of $a \in C$, respectively.

Definition 3. Let S = (U, C, D, V, I, n) be a NEDS, the error range of attribute a is defined as

$$n(a) = \Lambda e(a), \tag{6}$$

where

$$e(a) = \Delta \frac{\Sigma_{i=1}^{m} a(x_i)}{m}, \tag{7}$$

where $\Lambda \in [0, 1]$ is a user-specified parameter, and $a(x_i)$ is the i-th instance value of $a \in C$, $i \in [1, m]$, and m is the number of instance. The precision of $e(a)$ can be adjusted through Δ setting.

Obviously, if $\Lambda = 0$, a NEDS degrades to a decision system (DS). If $\Lambda = 1$ and $n(a)$ is a fixed value, a NEDS degrades to a decision system with error range (DS-ER) (see, e.g., [4]). Therefore NEDS is a generalization of DS and DS-ER.

We introduce how to deal with the abnormal value of measurement error. In applications, if the repeated measurement data satisfy:

$$|x_i - \overline{x}| > 3\sigma, (i = 1, 2, \ldots, N), \tag{8}$$

the x_i would be considered as an abnormal value and be rejected. Where x_i is the i-th measurement value, and $\overline{x}$ is the mean of all measurement values. This is the Pauta criterion of measurement error theory.

Now, we investigate the relationship between the limit of confidence interval and the standard deviation in the following proposition.

Proposition 4. *Let $-n(a)$ and $+n(a)$ be LCL and UCL, respectively, and Pr be the confidence level. We have the upper limit of confidence interval*

$$n(a) = 3\sigma, \tag{9}$$

where $Pr = 99.73\%$.

The value of exceed the confidence interval based on 99.73% confidence level is an abnormal error, which needs to be identified and removed from consideration. The standard normal distribution is a special case of the normal distribution. The limit of confidence interval is investigated in the following proposition.

Proposition 5. *Let $-n(a)$ and $+n(a)$ be LCL and UCL of standard normal distribution measurement errors, respectively. We have*

$$n(a) = 3. \tag{10}$$

Proof. The standard normal distribution is given by taking $\mu = 0$ mean and $\sigma^2 = 1$ in a general normal distribution. $n(a) = 3\sigma$, $n(a) > 0$. Therefore Equation (10) holds.

The adjusting factor Λ plays a key role in Definition 3. Related introduction is given by the following proposition.

Proposition 6. *Let $-n(a)$ and $+n(a)$ be LCL and UCL of $a \in C$, respectively. Confidence intervals are stated at the Pr confidence level, and $n(a) = 3\sigma$. According to Equation (3), we have*

$$Pr(-n(a) \leq x \leq n(a)) = F(n(a)) - (1 - F(n(a))). \tag{11}$$

According to Equation (3) and Proposition 6, if $2/3 \leq \Lambda < 1$, we have $2\sigma \leq n(a) < 3\sigma$, $95.45\% \leq Pr < 99.73\%$; if $1/3 < \Lambda < 2/3$, we have $\sigma < n(a) < 2\sigma$, $68.27\% < Pr < 95.45\%$, and if $0 < \Lambda \leq 1/3$, we have $0 < n(a) \leq \sigma$, $0\% < Pr \leq 68.27\%$.

One can adjust the size of the neighborhood through the Λ setting to meet different requirements.

2.3. Test-cost-independent decision system with normal distribution measurement errors

We introduce test costs to the data model. Now, we discuss the new model as follows:

Definition 7. A *test-cost-independent decision system with normal distribution measurement errors* (TCI-NEDS) S is the 7-tuple:

$$S = (U, C, D, V, I, n, c), \tag{12}$$

where U, C, D, V, I and n have the same meanings as in a NEDS, $c : C \to \mathbb{R}^+ \cup \{0\}$ is the test cost function. Test costs are independent of one another, that is, $c(B) = \sum_{a \in B} c(a)$ for any $B \subseteq C$.

Note that in this model, test costs are not applicable to decision attributes.

In order to processing and comparison, the values of conditional attributes are normalized from their value into a range from 0 to 1.

$$y = \begin{cases} (x - min)/(max - min) & \text{if } max \neq min; \\ 0.5 & \text{otherwise.} \end{cases} \tag{13}$$

where x is the initial value, y is the normalized value, and max and min are the maximal and minimal values of the attribute domain, respectively.

Table 1 presents a decision system of *Iris*, which conditional attributes are normalized values. Where $C = \{\text{SL, SW, PL, PW}\}$, $D = \{\text{Class}\}$, and $U = \{x_1, x_2, \ldots, x_{150}\}$.

Table 1: An example numerical value attribute decision table.

Patient	SL	SW	PL	PW	Class
x_1	0.23529	0.77273	0.14286	0.04762	setosa
x_2	0.29412	0.72727	0.11905	0.04762	setosa
x_3	0.35294	0.09091	0.38095	0.42857	versicolor
x_4	0.64706	0.31818	0.52381	0.52381	versicolor
x_5	0.41176	0.31818	0.50000	0.42857	versicolor
...	...	...	...	...	...
x_{149}	0.58824	0.54545	0.85714	1.00000	virginica
x_{150}	0.44118	0.27273	0.64286	0.71429	virginica

3. Covering-based rough set with normal distribution measurement errors

Rough set theory is a powerful tool for dealing with uncertain knowledge in information systems [31]. It has been successfully applied into feature selection [32, 33], rule extraction [34, 35, 36], uncertainty reasoning [37, 38], decision evaluation [35, 39, 40], granular computing [41, 42, 43, 44], etc. Recently,

covering-based rough set has attracted much research interest with significant achievements in both theory and application.

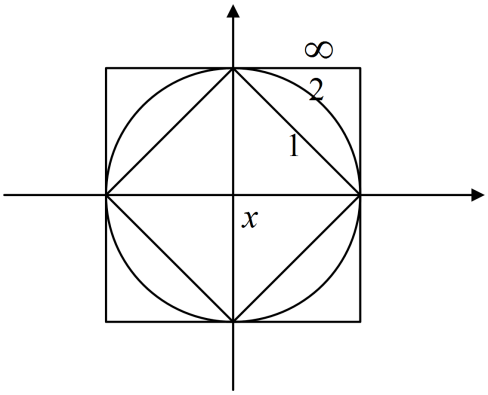

Figure 2: Conventional neighborhoods.

The concept of neighborhood (see, e.g., [45, 46, 47]) has been applied to define different types of covering-based rough set [13, 19, 23]. From the different viewpoints, a neighborhood is called an information granule, or a covering element. Figure 2 illustrates the neighborhoods of x in a two-dimension real space [25]. For this neighborhood rough set model, δ is a distance parameter and objects with a distance no further than δ are viewed as neighbors. In this approach, δ is a user-specified parameter. A new type of neighborhood is defined in [4], and Figure 3 illustrates this two-dimensional neighborhood. The size of the neighborhood depends on error ranges of tests, and more objects fall into the neighborhood of x_i for wider error ranges.

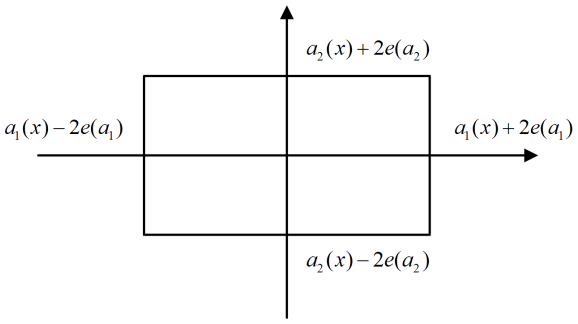

Figure 3: Two-dimensional neighborhood with error ranges.

In this section, we introduce normal distribution measurement errors to covering-based rough set. The new model is called covering-based rough set with normal distribution measurement errors. If all attributes are error free, the data in a neighborhood are equivalent to each other. In this case, the covering-based rough set model degenerates to the classical one. Therefore, the

covering-based rough set with NDME is a natural extension of classical rough set.

3.1. Covering-based rough set with normal distribution measurement errors

According to "3-sigma" rule, we present a new model considering both error distribution and confidence interval. With Definition 2, a new neighborhood is defined as follows.

Definition 8. Let $S = (U, C, D, V, I, n)$ be a NEDS. Given $x_i \in U$ and $B \subseteq C$, the neighborhood of x_i with respect to normal distribution measurement errors on test set B is defined as

$$n_B(x_i) = \{x \in U | \forall a \in B, |a(x) - a(x_i)| \leq 2n(a)\}, \tag{14}$$

where $n(a) = \Lambda e(a)$ is the error range based on confidence level of a. It represents the error value of a in $[-n(a), +n(a)]$.

Measurement errors with no more than a difference of $2n(a)$ should be viewed as the family of neighborhood granules. We explain why $n(a)$ instead of $e(a)$ was employed in Equation (14) as the maximal distance. Although the value of error is within a certain range, there are significant differences among confidence intervals. As mentioned earlier, "3-sigma" rule states that for a normal distribution, different proportion values lie within different standard deviations of the mean. Especially, the proportion is very close to 0 if data is more than three standard deviations from the mean. Therefore, measurement errors with no more than a difference of $n(a) = \Lambda e(a)$ should be viewed the family of neighborhood granules.

Sometimes we have a number of tests to obtain the same data item. Suppose some error ranges are known and others are unknown. The following proposition provides an estimation.

Proposition 9. *Let a_i and a_j be the measurement values for the same data item, $|a_j(x) - a_i(x)| \leq n'$ for any $x \in U$. We have*

$$e(a_j) \leq e(a_i) + n'/\Lambda. \tag{15}$$

Proof. Let the true value of $x \in U$ be $a^*(x)$ for $a \in B$. Due to the measurement error, $a^*(x) - \Lambda e(a_i) \leq a(x_i) \leq a^*(x) + \Lambda e(a_i)$. $a_j(x) \leq a_i(x) + n' \leq a^*(x) + (\Lambda e(a_i) + n')$; $a_j(x) \geq a_i(x) + n' \geq a^*(x) - (\Lambda e(a_i) + n')$. Therefore $e(a_j) \leq e(a_i) + n'/\Lambda$.

The shape of the neighborhoods is an ellipse for two-dimensional space. The two-dimensional block is depicted in Figure 4. Naturally, the size of the neighborhood depends on error ranges of tests and adjusting factor. Figure 5 shows the different sizes of neighborhood based on different adjusting factors.

Now we discuss some fundamental issues of rough set in the new model.

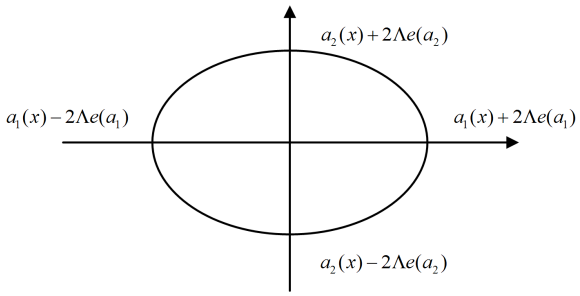

Figure 4: Two-dimensional neighborhood with NDME.

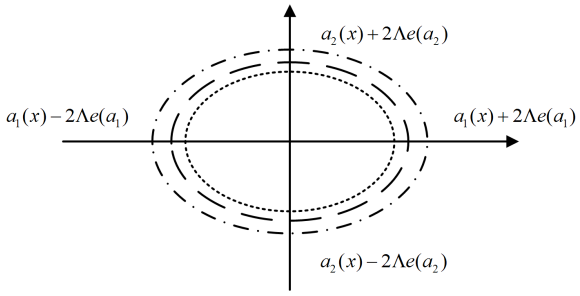

Figure 5: Two-dimensional neighborhood with NDME based on different adjusting factors.

Definition 10. Let $S = (U, C, D, V, I, n)$ be a NEDS, N_B be a neighborhood relation induced by $B \subseteq C$. We call $< U, N_B >$ a neighborhood approximation space. For any $X \subseteq U$, two subsets of objects, called lower and upper approximations of X in $< U, N_B >$, are defined as

$$\underline{N_B}(X) = \{x_i | x_i \in U \wedge n_B(x_i) \subseteq X\}; \tag{16}$$

$$\overline{N_B}(X) = \{x_i | x_i \in U \wedge n_B(x_i) \cap X \neq \emptyset\}; \tag{17}$$

$\forall X \subseteq U$, $\overline{N_B}(X) \supseteq X \supseteq \underline{N_B}(X)$. The boundary region of X in the approximation space is defined as

$$BN_B(X) = \overline{N_B}(X) - \underline{N_B}(X). \tag{18}$$

The positive region of D with respect to $B \subseteq C$ is defined as $POS_B(D) = \bigcup_{X \in U/D} \underline{N_B}(X)$ [48, 49].

3.2. Test-cost-sensitive attribute reduct problem

Attribute reduction is a successful technique to remove redundant data and facilitate the mining task. A number of definitions of relative reducts exist [25, 37, 50, 51] for different rough set models. In this section, we define test-cost-sensitive attribute reduction on the covering-based rough set model with NDME.

Definition 11. Let $S = (U, C, D, V, I, n)$ be a NEDS, $B \subseteq C$ and $x \in U$. Any $y \in n_B(x)$ is called an inconsistent object in $n_B(x)$ if $D(y) \neq D(x)$. The set of inconsistent objects in $n_B(x)$ is

$$ic_B(x) = \{y \in n_B(x) | D(y) \neq D(x)\}. \tag{19}$$

The number of inconsistent objects, namely $|ic_B(x)|$, is important in evaluating the characteristics of the neighborhood block. It also influences the quality of rule induced by the block.

From Definition 11 we know that given $B \subseteq C$, $x \in POS_C(D)$ if and only if $ic_R(x) = \emptyset$. Consequently, we have the following proposition, which can be employed as an alternative definition of a reduct.

Proposition 12. *Let $S = (U, C, D, V, I, n)$ be a NEDS. Any $R \subseteq C$ is a decision-relative reduct iff:*
1. $\forall x \in POS_C(D), ic_R(x) = \emptyset$, and
2. $\forall x \in R, \exists x \in POS_C(D), st. ic_{R-\{a\}}(x) \neq \emptyset$.

This proposition will help us in reduction algorithm designing, which as will be discussed in Section 4. Sometimes we are interested in minimal reduction or minimal test cost reduct (see, e.g., [9]). In this work, we focus on finding reducts with minimal test cost, that is, test-cost-sensitive attribute reducts. Since TCI-NEDS is a generalization of NEDS, concepts in the latter model are also applicable to the former one. We propose the following concept.

Definition 13. Let $Red(S)$ denote the set of all reducts of a TCI-NEDS $S = (U, C, D, V, I, n, c)$. Any $R \in Red(S)$ where $c(R) = min\{c(R') | R' \in Red(S)\}$ is called a *minimal test cost reduct*.

A minimal test cost reduct problem proposed in [9] can be redefined as follows. The problem of finding such a reduct is called the *minimal test cost reduct problem*.

Problem 14. *The minimal test cost reduct problem.*
Input: $S = (U, C, D, V, I, n, c)$;
Output: $B \subseteq C$;
Constraint: $POS_B(D) = POS_C(D)$;
Optimization objective: $min|c(B)|$.

Compared with the classical minimal reduction problem, there are several differences as follows. The first is the input, where the test costs and measurement errors are the external information. The second is the optimization objective, which is to minimize the test cost, instead of the number of features. We can adopt the addition-deletion strategy [14] to design our heuristic reduction algorithm.

3.3. Evaluation measures

Three evaluation measures are adopted to evaluate the performance of the proposed algorithm in order to dispel the influence of subjective and objective factors. We adopt the three measures proposed in [9] for this purpose. These are finding optimal factor (FOF), maximal exceeding factor (MEF), and average exceeding factor (AEF).

Let K be the number of experiments and k be the number of searched optimal reduct in the experiments. The finding optimal factor is defined as

$$op = \frac{k}{K}, \tag{20}$$

which is a both qualitative and quantitative measure.

Let R' be an optimal reduct and R be the searched reduct. The exceeding factor indicating the badness of a reduct is a quantitative measure, defined as

$$ef(R) = \frac{c(R) - c(R')}{c(R')}. \tag{21}$$

The maximal exceeding factor describes the worst case of an algorithm, defined as

$$\max_{1 \le i \le K} ef(R_i). \tag{22}$$

The average exceeding factor is defined as

$$\frac{\sum_{i=1}^{K} ef(R_i)}{K}, \tag{23}$$

which represents the whole performance of an algorithm.

4. Algorithm

Test-cost-sensitive attribute reduct problem is more complex than the traditional reduct problem [4]. Heuristic algorithms are needed to find sub-optimal reducts for large datasets. To evaluate the performance of a heuristic algorithm in terms of the quality of the solution, we should find an optimal reduct from all reducts. Hence, exhaustive algorithms are also needed.

In this section, we mainly present a heuristic algorithm and a competition approach to deal with the new problem. The exhaustive algorithm of [4] is adopted to find all reducts of datasets. It is based on backtracking where pruning techniques are crucial in reducing computation.

4.1. The δ-weighted heuristic reduction algorithm

To design a heuristic algorithm, we employ an algorithm framework which is similar to the one proposed in [9]. The algorithm follows the typical addition-deletion strategies [14], which is listed in Algorithm 1. It constructs a super-reduct, and then reduces it to obtain a reduct. The algorithm is essentially

different from the one in [9]. First, the input S is a TCI-NEDS instead of a test-cost-independent decision system (TCI-DS). Second, test results are numerical rather than nominal. The key code of this framework is listed in lines 5 and 7, and the attribute significance function is redefined to obtain respective algorithm. The efficiency of the δ-weighted heuristic reduction algorithm will be discussed in Section 5.4.

As previously mentioned, $|ic_B(x)|$ is useful in evaluating the quality of a neighborhood block. Now we propose the following concepts.

Definition 15. Let $S = (U, C, D, V, I, n)$ be a NEDS, $B \subseteq C$ and $x \in U$. The number of inconsistent objects in neighborhood $n_B(x)$ is $|ic_B(x)|$. The total number of such objects with respect to U is

$$nc_B(S) = \Sigma_{x \in U} |ic_B(x)|, \tag{24}$$

and with respect to the positive region is

$$pc_B(S) = \Sigma_{x \in POS_C(D)} |ic_B(x)|. \tag{25}$$

Finally, we propose a δ-weighted heuristic information function:

$$f(B, a_i, c(a_i)) = \Phi + \delta \frac{\Phi}{c(a_i)}, \tag{26}$$

where $\Phi = pc_B(S) - pc_{B \cup \{a_i\}}(S)$ is necessary and indispensable, and it plays a dominant role in the heuristic information. Where $c(a_i)$ is the test cost of a_i, and $\delta \geq 0$ is a user-specified parameter. If $\delta = 0$, test costs are essentially not considered. If $\delta > 0$, tests with lower cost have bigger significance. Different δ settings can adjust the significance of test cost.

4.2. The competition approach

The competition approach has been discussed in [9] to obtain better results with more run-time. In the new environment, it is still valid because there is no universally optimal δ. In this approach, reducts complete against each other with only one winner, that is a reduct with minimal test cost, which can be obtained using $\delta \in L$.

$$C_L = min_{\delta \in L} c(R_\delta), \tag{27}$$

where R_δ is the reduct obtained by Algorithm 1 using the heuristic information, with L the set of user-specified δ values.

This approach requires more run-time because the algorithm run $|L|$ times with different δ values. Since the heuristic algorithm is fast, it is acceptable for relatively small $|L|$. The results will be shown in Section 5.3. This simple approach can enhance the quality of the result significantly.

Algorithm 1 An addition-deletion test-cost-sensitive reduction algorithm.

Input: $(U, C, D, \{V_a\}, \{I_a\}, n, c)$
Output: A reduct with minimal test cost
Method:

```
1: B = ∅;
   //Addition
2: CA = C;
3: while (POS_B(D) ≠ POS_C(D)) do
4:    for each a ∈ CA do
5:       Compute f(B, a, c);
6:    end for
7:    Select a' with the maximal f(B, a', c);
8:    B = B ∪ {a'}; CA = CA − {a'};
9: end while
   //Deletion
10: CD = B;
11: while (CD ≠ ∅) do
12:    CD = CD − {a'};
13:    if (POS_{B−{a'}}(D) = POS_B(D)) then
14:       B = B − {a'};
15:    end if
16: end while
17: return B;
```

5. Experiments

5.1. Data generation

Most datasets from the UCI library [52] have no intrinsic measurement errors and test costs. In order to help to study the performance of the reduction algorithm, we will create some data for experimentations. In this way, different parameters can be specified and data satisfying normal distributions can be employed. Unlike in simpler models, data should not be randomly generated, but meet certain constraints. For example, measurement errors satisfy normal distribution and Pauta criterion. For the same data item, tests with narrower error ranges should be more expensive. In this section, we will discuss both the process and substantial settings of data generation. Constraints mentioned above are met in this process.

Step 1. We choose ten datasets from the UCI Repository of Machine Learning Databases, as listed in Table 2. Each dataset should contain exactly one decision attribute, and have no missing value. To make the data easier to handle, data items are normalized from their value into a range from 0 to 1. Missing values are directly set to 0.5.

Step 2. We produce the number of additional tests for one particular data item. We use the uniform distribution generator [9] to generate the random

Table 2: Database information.

No.	Name	Domain	$\|U\|$	$\|C\|$	$\|C'\|$	$D = d$
1	Iris	zoology	150	4	4	class
2	Glass	manufacture	214	9	13	type
3	Wine	agriculture	178	13	21	class
4	Wpbc	clinic	198	33	65	outcome
5	Wdbc	clinic	569	30	58	diagnosis
6	Credit	commerce	690	15	23	class
7	Image	graphics	210	19	30	class
8	Iono	physics	351	34	68	class
9	Liver	clinic	345	6	8	selector
10	Diab	clinic	768	8	12	class

integers in the range [0, k]. That is, we have 1 to (k + 1) measurement methods to obtain values for each data item; k is set to less than 5 in our experiments. The number of tests for our experiments is $|C'|$ in Table 2.

Table 3: Generated error ranges for different databases.

Datasets	Minimal	Maximal	Average
Iris	0.0042	0.0044	0.0043
Glass	0.0005	0.0059	0.0030
Wine	0.0031	0.0053	0.0040
Wpbc	0.0011	0.0056	0.0031
Wdbc	0.0006	0.0040	0.0023
Credit	0.0001	0.0056	0.0022
Image	0.0001	0.0065	0.0026
Iono	0.0045	0.0087	0.0061
Liver	0.0011	0.0065	0.0029
Diab	0.0009	0.0059	0.0031

Step 3. We produce the $e(a)$ for each original test according to Equation (7). The $e(a)$ is computed according to the value of databases without any subjectivity. Three kinds of error ranges of different databases are shown in Table 3. These error ranges are maximal, minimal and average error ranges of all attributes, respectively. The precision of $e(a)$ can be adjusted through Δ setting, and we set Δ to be 0.01 in our experiments.

Step 4. We produce "new" data subject to error ranges. Let a_1 be the original test, according to Proposition 9, we can add a random number in $[-(i-1)n(a), (i-1)n(a)]$ to $a_1(x)$ to produce $a_i(x)$, where $x \in U$. The number is generated as follows.

Let x_1 and x_2 be uniformly distributed on (0, 1), then

$$y_1(x_1, x_2) = \sqrt{-2 \ln x_1} cos(2\pi x_2) \tag{28}$$

is a random number which has a normal distribution with $\mu = 0$ mean and $\sigma^2 = 1$. From Proposition 4 we know the $n(a) = 3\sigma$, and $\sigma = \frac{1}{3}n(a)$.

Since we need a random number in $[-n(a), +n(a)]$, we let

$$y(n(a), x_1, x_2) = \frac{1}{3} y_1(x_1, x_2) n(a). \tag{29}$$

Finally

$$NDME = \begin{cases} -n(a) & \text{if } y < -n(a); \\ n(a) & \text{if } y > n(a); \\ y(n(a), x_1, x_2) & \text{otherwise.} \end{cases} \tag{30}$$

is a random number which has a normal distribution with $\mu = 0$ mean and $\sigma = \frac{1}{3}n(a)$. According to Definition 8, a_i is the new test with error range $\pm i * n(a)$.

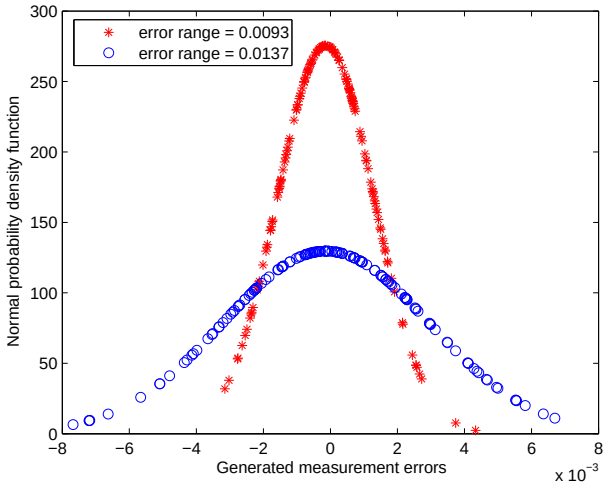

Figure 6: Normal distribution measurement errors with different error ranges.

The generated NDME with different error ranges are shown in Figure 6. The generated NDME of different databases are shown in Figure 7.

Step 5. We produce test costs, which are always represented by positive integers. Let a_1 be the original test and a_1 be the last test for one particular data item. $c(a_l)$ is set to a random number in [1, 100] subject to the uniform distribution. $c(a_i)$ where $1 \leq i < l$ is set to $2 \times c(a_{i+1})$. This setting guarantees that tests with narrower error ranges are more expensive.

A dataset generated by this approach is listed in Table 4. SL stands for sepal length, SW stands for sepal width, PL stands for petal length, and PW

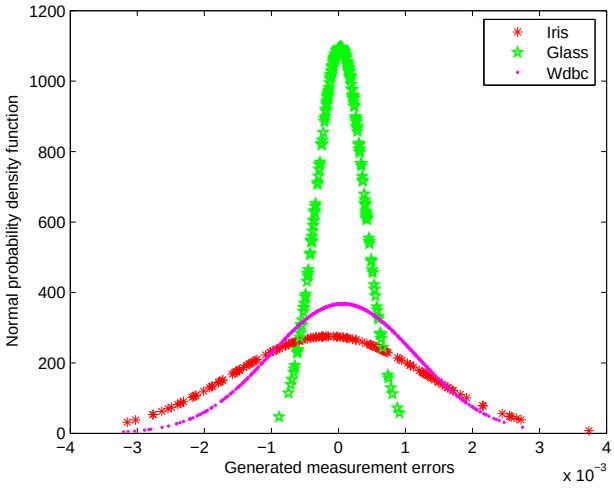

Figure 7: Normal distribution measurement errors of Iris, Glass, and Wdbc.

Table 4: A generated measurement error vector and a generated test cost vector(Iris).

a	SL	SL-1	SW	PL	PL-1	PL-2	PW
Original test	True	False	True	True	False	False	True
$e(a)$	0.0043	0.0086	0.0041	0.0041	0.0082	0.0123	0.0041
$c(a)$	28	14	81	376	188	94	91

stands for petal width. 1 or 2 after SL and PL indicate different revisions of the original data. There is only one method to obtain SW and PW.

5.2. Effectiveness of the heuristic algorithm

Let $\delta = 2, 3, 4, \ldots, 9$. The algorithm runs 800 times with different test cost settings and different δ setting on all datasets. Figures 8 and 9 show the results of finding optimal factors. For different settings of δ, the performance of the algorithm is completely different, that is, the test cost plays a key role in this heuristic algorithm. Data for $\delta = 0$ are not included in the experiment results because respective results are incomparable to others.

Figures 10 and 11 show the results of maximal exceeding factors, which provide the worst case of the algorithm, and they should be viewed as a statistical measure. Figures 12 and 13 show the average exceeding factors. These display the overall performance of the algorithm from a statistical perspective.

From the results we observe that the quality of the results varies for different datasets. It is related to the dataset itself because the error range and heuristic information are all computed according to the values of dataset. Then the

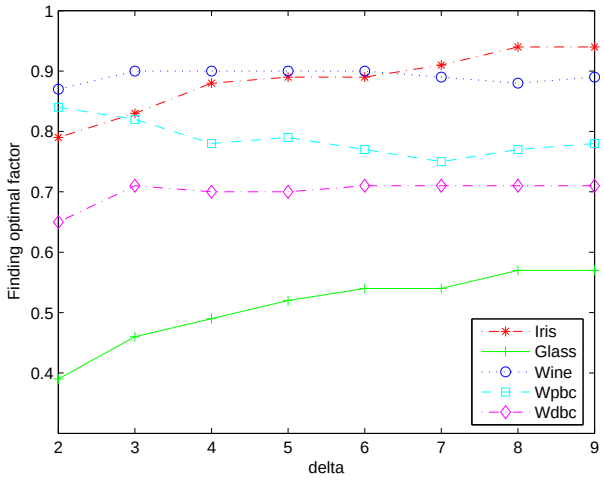

Figure 8: Finding optimal factor (datasets 1-5).

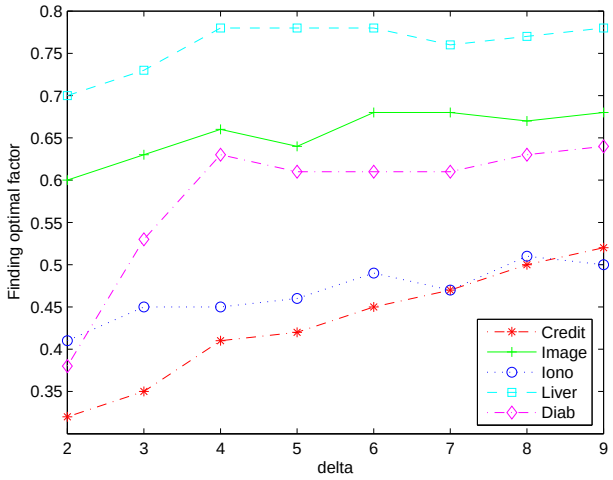

Figure 9: Finding optimal factor (datasets 6-10).

average exceeding factor is less than 0.3 in most cases. In other words, the results are acceptable. Although the results are generally acceptable, the performance of the algorithm should be improved. Section 5.3 will address this issue further.

5.3. Comparison of three approaches

Now we compare the performance of the proposed algorithm through three approaches mentioned in Section 4. The first approach, called the non-weighting

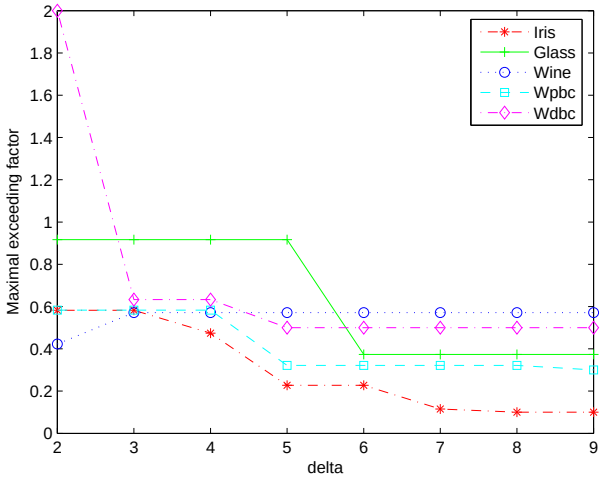

Figure 10: Maximal exceeding factor (datasets 1-5).

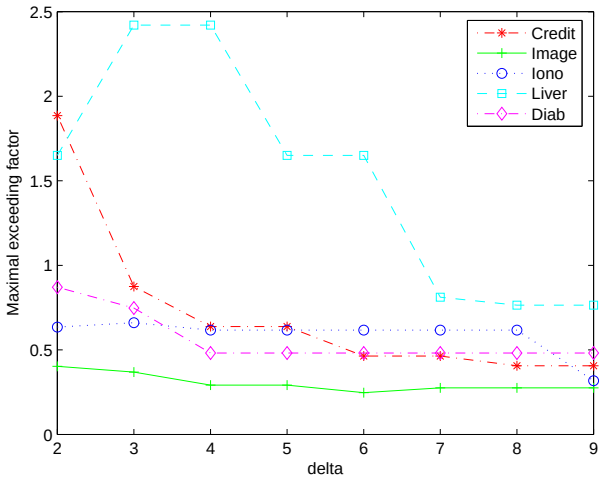

Figure 11: Maximal exceeding factor (datasets 6-10).

approach, is implemented by setting $\delta = 0$. This approach is the only one without taking into account test costs. The second approach, called the best δ approach, is to choose the best δ value as depicted in Figures 8 through 13. The third approach is the competition approach discussed in Section 4.2. All three are based on Algorithm 1 and the same databases.

Table 5 lists results for all three approaches. From Table 5, we observe the following results:

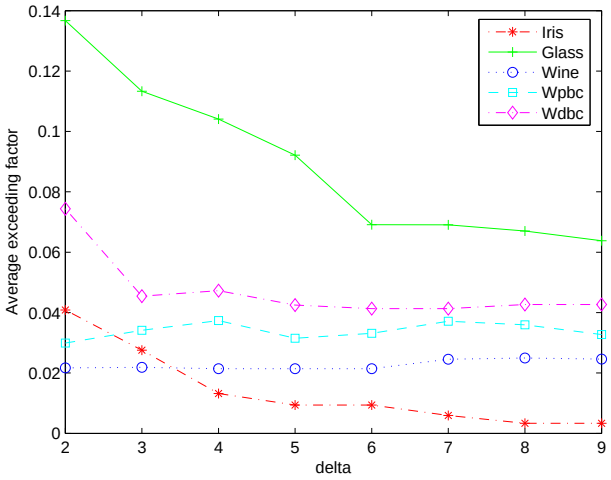

Figure 12: Average exceeding factor (datasets 1-5).

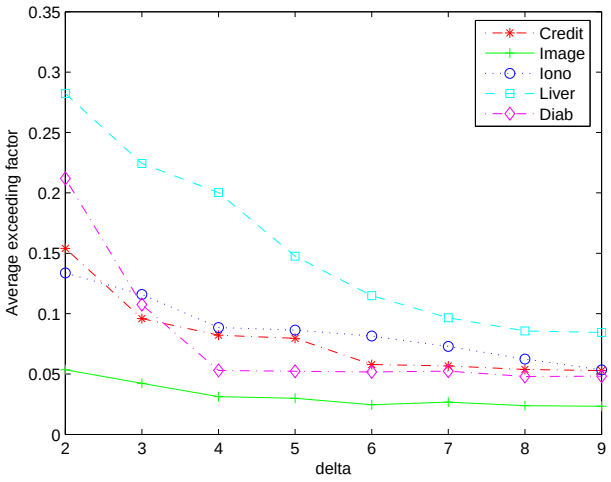

Figure 13: Average exceeding factor (datasets 6-10).

(1) The non-weighting approach almost does not find the optimal reduct. Therefore without considering test costs is not suitable for the minimal test cost reduct problem.
(2) In most cases, the best δ approach obtains good results. However, we have no idea how to obtain the best value of δ in real applications.
(3) The competition approach significantly improves the quality of results with more run-time, which is acceptable for relatively small number of δ.

Table 5: Results for $\delta = 0, \delta$ with the optimal setting, and δ with a number of choices.

Dataset	FOF			MEF			AEF		
	$\delta = 0$	$\delta = \delta^*$	$\delta \in L$	$\delta = 0$	$\delta = \delta^*$	$\delta \in L$	$\delta = 0$	$\delta = \delta^*$	$\delta \in L$
Iris	0.170	0.940	0.940	2.000	0.100	0.100	0.360	0.003	0.003
Glass	0.090	0.570	0.640	3.220	0.374	0.374	0.700	0.064	0.049
Wine	0.000	0.900	0.940	19.44	0.423	0.423	4.464	0.021	0.014
Wpbc	0.000	0.840	0.880	45.67	0.300	0.250	14.50	0.033	0.017
Wdbc	0.000	0.710	0.760	93.20	0.500	0.500	14.61	0.041	0.037
Credit	0.000	0.520	0.550	2.188	0.317	0.310	1.095	0.053	0.049
Image	0.000	0.680	0.790	31.43	0.406	0.269	5.417	0.053	0.032
Iono	0.000	0.500	0.630	46.60	0.765	0.544	10.28	0.084	0.054
Liver	0.040	0.780	0.910	4.125	0.275	0.181	0.921	0.023	0.008
Diab	0.000	0.640	0.700	3.788	0.481	0.481	1.278	0.048	0.033

5.4. Comparison with existing algorithm

Compared with an existing model [4], the major improvement is introduced in this section.

First, the NDME was considered to data model, and covering-based rough set based on NDME has been proposed. In most cases, the measurement errors satisfy normal distribution instead of uniform distribution; hence, this new model has wider application areas.

Second, comparing with the fix error range of different databases from [4], the proposed error ranges are adaptively generated according to the database values. Table 3 shows the generated error ranges for different databases. The error ranges for different attributes of the same database are completely different. For example, the maximal error range of Wdbc is 0.0040, and the minimal one is 0.0006.

Third, a δ-weighted heuristic algorithm is developed to deal with the minimal test cost reduct problem. Our algorithm is compared with the λ-weighted algorithm [4] from effectiveness and efficiency. Since two different algorithms have different parameters, we compare the results of the competition approach on ten datasets. Figure 14 shows competition approach results of two algorithms. From the results we observe that

(1). On Wpbc and Iono datasets, two algorithms have same performance.

(2). λ-weighted algorithm has better performance than our algorithm on Iris, Class and Credit datasets.

(3). However, our algorithm performs better than the λ-weighted algorithm on five datasets.

The efficiency comparison between the δ-weighted algorithm and λ-weighted one is depicted in Figure 15. From the results we note that our algorithm has an improvement in terms of run-time. Figure 16 shows the efficiency ratios of the δ-weighted algorithm and the λ-weighted algorithm.

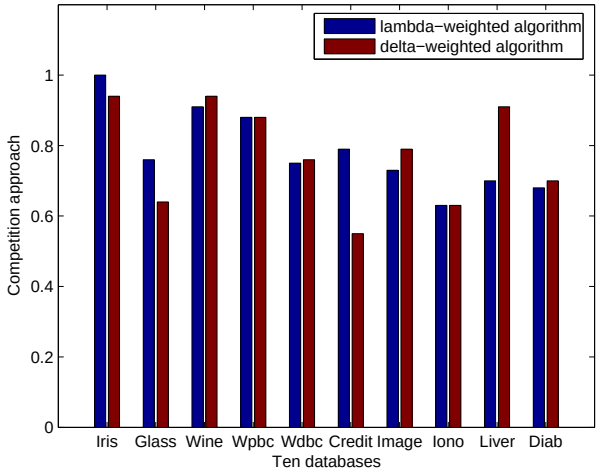

Figure 14: Competition approach results of two algorithms.

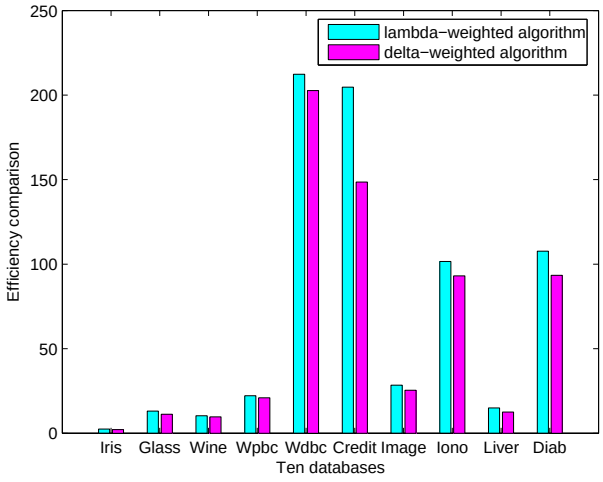

Figure 15: Efficiency comparison.

6. Conclusion

In rough set model, measurement errors and test costs are all intrinsic to data. In this paper, we built a new covering-based rough set model considering measurement errors and test costs at four levels:

1. At the data model level, a new data model with NDME and test cost was proposed. This model has more application areas because the measure-

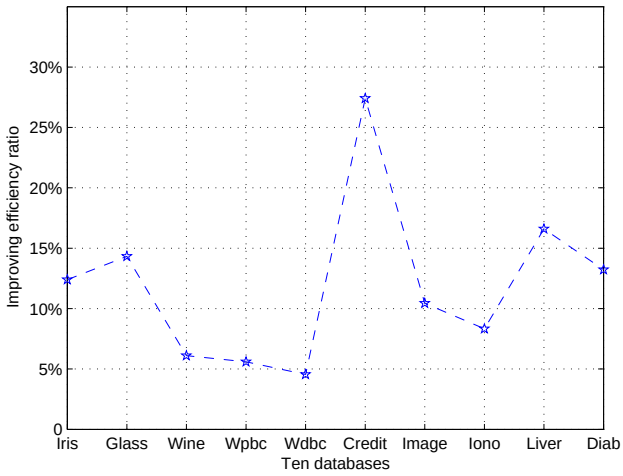

Figure 16: Improving efficiency ratio.

ment errors have certain universality.

2. At the computational model level, we introduced a covering-based rough set with NDME. This model is generally more complex than that presented in this field.
3. At the problem level, a minimal test cost reduct problem based on the new model was redefined.
4. At the algorithm level, a δ-weighted heuristic algorithm was developed to deal with this reduct problem. Experimental results indicate the effectiveness and efficiency of the algorithm.

In summary, the data model based on normal distribution measurement errors has wide application scope. This study suggests new research trends of covering-based rough set and cost-sensitive learning.

Acknowledgment

This work is in part supported by the Fujian Province Foundation of Higher Education (No. JK2012028), the National Nature Science Foundation of China (No. 61170128), and the Natural Science Foundation of Fujian Province, China (Nos. 2011J01374, 2012J01294).

References

[1] S. Bell, *A beginner's guide to uncertainty of measurement.* National Physical Laboratory, 2001.

[2] M. Chau, R. Cheng, B. Kao, and J. Ng, "Uncertain data mining: an example in clustering location data," in *Proceedings of Advances in Knowledge Discovery and Data Mining*, ser. LNCS, vol. 3918, 2006, pp. 199–204.

[3] C. C. Aggarwal, "On density based transforms for uncertain data mining," in *Proceedings of IEEE 23rd International Conference on Data Engineering*, 2007, pp. 866–875.

[4] F. Min and W. Zhu, "Attribute reduction of data with error ranges and test costs," *Information Sciences*, vol. 211, pp. 48–67, 2012.

[5] M. Pazzani, C. Merz, P. M. K. Ali, T. Hume, and C. Brunk, "Reducing misclassification costs," in *Proceedings of the 11th International Conference of Machine Learning (ICML)*. Morgan Kaufmann, 1994, pp. 217–225.

[6] W. Fan, S. Stolfo, J. Zhang, and P. Chan, "Adacost: Misclassification cost-sensitive boosting," in *The 16th International Conference on Machine Learning (ICML)*, 1999, pp. 97–105.

[7] Z. Zhou and X. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.

[8] F. Min and Q. Liu, "A hierarchical model for test-cost-sensitive decision systems," *Information Sciences*, vol. 179, pp. 2442–2452, 2009.

[9] F. Min, H. He, Y. Qian, and W. Zhu, "Test-cost-sensitive attribute reduction," *Information Sciences*, vol. 181, pp. 4928–4942, 2011.

[10] H. Zhao, F. Min, and W. Zhu, "Test-cost-sensitive attribute reduction based on neighborhood rough set," in *Proceedings of the 2011 IEEE International Conference on Granular Computing*, 2011, pp. 802–806.

[11] Y. Yao and Y. Zhao, "Attribute reduction in decision-theoretic rough set models," *Information Sciences*, vol. 178, no. 17, pp. 3356–3373, 2008.

[12] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artificial Intelligence*, vol. 151, pp. 155–176, 2003.

[13] W. Zhu and F. Wang, "Reduction and axiomization of covering generalized rough sets," *Information Sciences*, vol. 152, no. 1, pp. 217–230, 2003.

[14] Y. Y. Yao, Y. Zhao, and J. Wang, "On reduct construction algorithms," in *Proceedings of Rough Set and Knowledge Technology*, ser. LNAI, vol. 4062, 2006, pp. 297–304.

[15] H. X. Li, X. Z. Zhou, J. B. Zhao, and D. Liu, "Attribute reduction in decision-theoretic rough set model: A further investigation," in *Proceedings of Rough Sets and Knowledge Technology*, ser. LNCS, vol. 6954, 2011, pp. 466–475.

[16] X. Y. Jia, W. H. Liao, Z. M. Tang, and L. Shang, "Minimum cost attribute reduction in decision-theoretic rough set models," *Information Sciences, doi: j.ins.2012.07.010*, 2012.

[17] W. Zhu and F. Wang, "Covering based granular computing for conflict analysis," *Intelligence and Security Informatics*, pp. 566–571, 2006.

[18] W. Zhu, "Topological approaches to covering rough sets," *Information Sciences*, vol. 177, no. 6, pp. 1499–1508, 2007.

[19] ——, "Generalized rough sets based on relations," *Information Sciences*, vol. 177, no. 22, pp. 4997–5011, 2007.

[20] W. Zhu and F. Wang, "On three types of covering-based rough sets," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 19, no. 8, pp. 1131–1144, 2007.

[21] W. Zhu, "Relationship among basic concepts in covering-based rough sets," *Information Sciences*, vol. 179, no. 14, pp. 2478–2486, 2009.

[22] ——, "Relationship between generalized rough sets based on binary relation and covering," *Information Sciences*, vol. 179, no. 3, pp. 210–225, 2009.

[23] L. Ma, "On some types of neighborhood-related covering rough sets," *International Journal of Approximate Reasoning*, 2012.

[24] Q. Hu, D. Yu, and Z. Xie, "Numerical attribute reduction based on neighborhood granulation and rough approximation (in chinese)," *Journal of Software*, vol. 19, no. 3, pp. 640–649, March 2008.

[25] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.

[26] Q. Hu, W. Pedrycz, D. Yu, and J. Lang, "Selecting discrete and continuous features based on neighborhood decision error minimization," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 40, no. 1, pp. 37–50, 2010.

[27] H. Li, M. Wang, X. Zhou, and J. Zhao, "An interval set model for learning rules from incomplete information table," *International Journal of Approximate Reasoning*, vol. 53, pp. 24–37, 2012.

[28] W. Wei, J. Liang, and Y. Qian, "A comparative study of rough sets for hybrid data," *Information Sciences*, vol. 190, pp. 1–16, 2012.

[29] F. Min, W. Zhu, H. Zhao, and G. Pan, "Coser: Cost-senstive rough sets, http://grc.fjzs.edu.cn/˜fmin/coser/," 2011.

[30] "Wikipedia." [Online]. Available: http://www.wikipedia.org

[31] P. Zhu, "Covering rough sets based on neighborhoods: An approach without using neighborhoods," *International Journal of Approximate Reasoning*, vol. 52, pp. 461–472, 2011.

[32] Y. Chen, D. Miao, and R. Wang, "A rough set approach to feature selection based on ant colony optimization," *Pattern Recognition Letters*, vol. 31, no. 3, pp. 226–233, 2010.

[33] Q. Hu, S. An, and D. Yu, "Soft fuzzy rough sets for robust feature evaluation and selection," *Information Sciences*, vol. 180, no. 22, pp. 4384–4400, 2010.

[34] Y. Du, Q. Hu, P. Zhu, and P. Ma, "Rule learning for classification based on neighborhood covering reduction," *Information Sciences*, 2011.

[35] Y. Qian, J. Liang, and C. Dang, "Converse approximation and rule extraction from decision tables in rough set theory," *Computers & Mathematics with Applications*, vol. 55, no. 8, pp. 1754–1765, 2008.

[36] X. Yang, D. Yu, J. Yang, and X. Song, "Difference relation-based rough set and negative rules in incomplete information system," *International Journal Of Uncertainty, Fuzziness, And Knowledge-based Systems*, vol. 17, no. 5, p. 649, 2009.

[37] Z. Pawalk, "Rough sets: theoretical aspects of reasoning about data," 1991.

[38] Z. Pawlak and A. Skowron, "Rough sets and boolean reasoning," *Information Sciences*, vol. 177, no. 1, pp. 41–73, 2007.

[39] Y. Qian, J. Liang, D. Li, H. Zhang, and C. Dang, "Measures for evaluating the decision performance of a decision table in rough set theory," *Information Sciences*, vol. 178, no. 1, pp. 181–202, 2008.

[40] Y. Leung, "Theory and applications of granular labelled partitions in multi-scale decision tables," *Information Sciences*, no. 181, pp. 3878–3897, 2011.

[41] R. Barot and T. Lin, "Granular computing on covering from the aspects of knowledge theory," in *Fuzzy Information Processing Society, 2008. NAFIPS 2008. Annual Meeting of the North American*. IEEE, 2008, pp. 1–5.

[42] S. Calegari and D. Ciucci, "Granular computing applied to ontologies," *International journal of approximate reasoning*, vol. 51, no. 4, pp. 391–409, 2010.

[43] T. Lin, "Granular computing: practices, theories, and future directions," *Encyclopedia of Complexity and Systems Science*, vol. 2009, pp. 4339–4355, 2009.

[44] L. Zadeh, "Fuzzy sets and information granularity," *Advances in fuzzy set theory and applications*, vol. 11, pp. 3–18, 1979.

[45] T. Y. Lin, "Granular computing on binary relations-analysis of conflict and chinese wall security policy," in *Proceedings of Rough Sets and Current Trends in Computing*, ser. LNAI, vol. 2475, 2002, pp. 296–299.

[46] A. Bargiela and W. Pedrycz, *Granular Computing: An Introduction*. Kluwer Academic Publishers, Boston, 2002.

[47] T. Y. Lin, "Granular computing - structures, representations, and applications," in *Lecture Notes in Artificial Intelligence*, vol. 2639, 2003, pp. 16–24.

[48] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, pp. 341–356, 1982.

[49] ——, *Rough sets: theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Boston, 1991.

[50] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, "Positive approximation: An accelerator for attribute reduction in rough set theory," *Artificial Intelligence*, vol. 174, no. 9-10, pp. 597–618, 2010.

[51] J. G. Bazan and A. Skowron, "Dynamic reducts as a tool for extracting laws from decision tables," in *Proceedings of the 8th International Symposium on Methodologies for Intelligent Systems*, 1994, pp. 346–355.

[52] C. L. Blake and C. J. Merz, "UCI repository of machine learning databases, http://www.ics.uci.edu/˜mlearn/mlrepository.html," 1998.